



TITLE:

1分子時系列情報から我々は何を学
び取ることができるか?(非線形科
学と統計科学の対話,研究会報告)

AUTHOR(S):

小松崎, 民樹

CITATION:

小松崎, 民樹. 1分子時系列情報から我々は何を学
び取ることができるか?(非線形科学と統計科学の対話,研究会報告). 物性研究 2008, 91(2):
121-130

ISSUE DATE:

2008-11-20

URL:

<http://hdl.handle.net/2433/142698>

RIGHT:

1 分子時系列情報から我々は何を学ぶことができるか？

北海道大学電子科学研究所

分子生命数理研究分野

小松崎民樹*

*tamiki@es.hokudai.ac.jp

1. はじめに

生体機能とは「外界からの刺激に対する応答として始まる一連の構造変化とそれに伴う化学反応」からなるダイナミックスの連関の産物であり、生命システムとは、常に外界に晒されながら、その生体機能を頑健に有する。そこでは、揺らぎを伴うメソスコピックな細胞レベルの非平衡性と分子の個別性が重要となるミクロスコピックな化学反応や分子の構造変化がダイナミカルに結合する。近年の計測技術の飛躍的な進展により、これまで集団平均に埋もれていた生体分子、それらの複合体、および細胞レベルにおけるダイナミックスを1分子レベルで追跡することが可能となり、「観測」の在り方が大きな変貌を遂げようとしている。すなわち、サブミリ秒程度の時間分解能で、1分子レベルの大規模構造変形や細胞の分化の経時変化を直接観測することが可能になってきた[1]。生命システムにおいて重要な事実は「分子レベルのイベントがマクロレベルでの機能を創出する」ことにある。すなわち、ミクロとマクロの関係を独立に取り扱う隷属原理に依拠する考えでは、生命システムは理解不可能である。このような生命システムに対してどのような理論的なアプローチがあり得るのだろうか？

生命システムを理解するための理論的なアプローチには、大別して、トップダウン的な構成論的アプローチとボトムアップ的な還元論的アプローチが存在する。前者は生命現象に関する(研究者の)確信的な先入観が介在することは避けられないし、その結果、自然と乖離したモデルに陥る可能性は常に存在する。一方で、後者は、分子論に固執するあまりに、枚挙主義に陥る危険性が高い。我々はトップダウン的にモデルを“構成する”のではなく、実際に観測される1分子時系列情報から「状態」、「状態間遷移ネットワーク構造」を評価し、相当するモデルをできる限り自然な形で“抽出”し、そのモデルに立脚して「システムと階層性」、「ミクロとマクロの階層間の情報伝達」、「多様性、適応性」を考察できるかを問う立場を取る。

仮にマイクロ秒以下の高時間分解能で1分子を追跡することが可能になるとすると、どのような新現象が発見され、更にはそれに伴ってどのような指導原理が新たに創出されるであろうか？不連続な“階段”としてのみ観測されてきた過渡的な遷移過程の直接観測が可能になるであろう。その場合には、遷移の非統計性 [2]が実験的に論じることが可能になるかもしれない。また、階層的なエネルギー地形 [3]の存在、階層間の非隷属性、更には、「自由エネルギー地形」概念の妥当性の再考、など興味深い研究対象が現れることが期待される。また、観測する時間分解能をマイクロ秒以下に上げることは、従来の時間分解能に比べて熱揺らぎの時間スケールに近づく。マイクロ秒の時間領域で起こる蛋白質ダイナミックスと熱揺らぎの拮抗関係、時間・空間スケールの異なる階層間の情報伝達・相関性など様々な魅力的なテーマが待ち構えている。1分子時系列から構成される記憶地形(Memory Landscape)[4]などの時系列解析手法が提案されているが、自由エネルギー地形ならびに状態遷移ネットワーク構造との関係が不明瞭であり、上記の展望に応えるだけの解析理論は現時点では未開拓といえる。

本稿では、我々が最近開発した局所平衡仮定に依拠しない計算力学[5]に基づく時系列解析理論を紹介し、時系列情報を通して観測されるダイナミックス描像を概観し、本解析理論のもつ将来性を展望する。通常、観測量(例えば、FRET 効率)から評価したい物理量(例えば、色素分子間距離)に

変換する必要がある(例えば、フィッシャー情報基準を用いた文献[6]などを参照していただきたい)が、ここでは、前処理後の時系列データから如何にして背後に存在する系の状態遷移ネットワークを構成できるか、また、そこから生体分子系に関するどのような動態描像が得られるかを問う。

2. 時系列データの背後に潜む状態遷移ネットワーク

どのように 1 分子時系列データから「状態」、そして「状態の遷移」を定義するのが自然であろうか？時系列データの観測値に対するヒストグラムを正規分布の線形結合でフィットする状態推定が古典的にはよく用いられてきたが、局所平衡の成立を陰に前提としており、一般に「正規」分布の重ね合わせで表わされる保証はどこにもない。予め、分布関数の形状、状態数や系についての性質(局所平衡等)を規定するのではなく、状態および状態遷移を“時系列情報からできるだけ自然な形で学びとる”方法論が必要となる。最近、我々は情報理論に依拠する計算力学[5]と Wavelet 多重解像度分解を併用し、背後に存在する階層的な状態遷移ネットワークを抽出する時系列解析手法を開発した [7]。以下、計算力学を 2-1, 2-2, 2-3 章で概説し、2-4 章で我々の解析手法を紹介することにする。

2-1. 時系列情報から「状態」を定義する

通常、時系列データはなんらかの実験誤差を包んでいるため、時系列データを離散化(シンボル化)することが自然である。以下ではシンボル化された時系列 X を取り扱うことにする。計算力学では、以下のようにして因果性に着目してシンボル列から“状態”を定義する。例えば、連続する v 個の時刻 $t_i, t_{i+1}, \dots, t_{i+v-1}$ における X 値から成るシンボル列 $s[n]$ ($=X(t_i)X(t_{i+1})\cdots X(t_{i+v-1})$)を考える(ここで変数 n は、一般に複数存在する(であろう)シンボル列パターンの n 番目を意味する)。このとき、あるシンボル列 $s[n]$ に継続して発生する“未来の”シンボル列 $s[j]$ ($=X(t_{i+v})X(t_{i+v+1})\cdots X(t_{i+2v-1})$)に対して、 $P(j|n)$ を $s[n]$ の直後に $s[j]$ が時系列情報 X のなかから見出される条件つき確率とする。計算力学では、任意の j に対して、 $P(j|n) \approx P(j|m)$ が成り立つとき、“過去”配列 $s[n]$ と $s[m]$ は同じ「状態」に帰属されるものと考え。そして、このようにして規定された「状態」を因果状態 S_i (Causal State)と呼び、(一般に幾つかの異なる時系列断片の集合から成る)因果状態は、定義から状態毎に固有の遷移確率分布を持つことになる。長さ v の時系列断片はすべて、ある因果状態に属することになるので、異なる因果状態間を繋ぐ状態遷移ネットワーク(ϵ マシンと呼ぶ)を構成することができる。このとき、因果状態 S_i から S_j への遷移の出現確率は始点の S_i にだけ依存するように、因果状態を規定するシンボル長 v を決定する。具体的には、以下に述べる因果状態分布の平均情報量(=統計的複雑さ Statistical Complexity)がシンボル長に依存しなくなる長さを最適な v とする。因果状態を(内部構造をもった)準安定状態とみなせば、1 分子時系列情報から離散的マルコフ過程を状態空間に構成したことに対応するため、関連する種々の物理量を解析的に評価することができる。

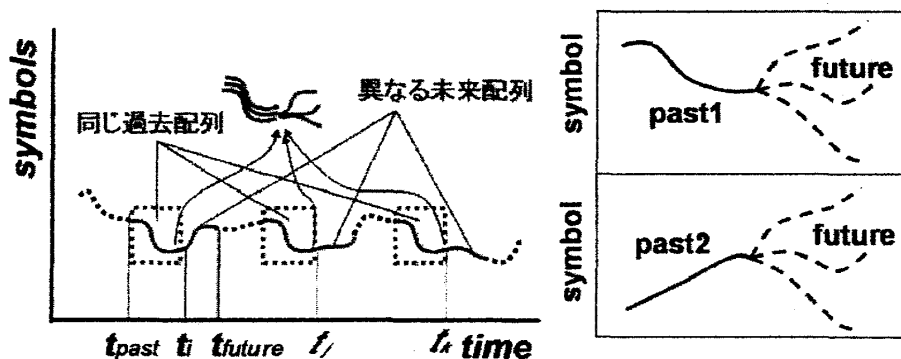


図 1 因果状態(Causal State)の構成: (左図)同じ“過去”配列 $s[n]$ から続いて生起する“未来”配列 $s[j]$ は一般に異なり、各々異なる生起確率 $P(j|n)$ (条件付き遷移確率) が付与される。(右図)異なる過去配列 $\text{past } 1, \text{past } 2$ に対して、任意の j に対して、 $P(j|\text{past } 1) \approx P(j|\text{past } 2)$ が成り立つとき、 $\text{past } 1$ と $\text{past } 2$ は同じ因果状態を構成する

2-2. 状態間遷移ネットワークの複雑さとεマシンの数学的性質

因果状態分布の平均の情報量(因果状態の滞在確率に対するShannon エントロピー)を統計的複雑さ(Statistical Complexity) C_μ [5]と呼ぶ。因果状態の総数を N 、因果状態 S_i の存在確率を $P(S_i)$ とすると、

$$C_{\mu} = -\sum_{I=1}^N P(S_I) \log_2 P(S_I)$$

と表される。この他、状態数のみを反映したTopological Complexity $C_0 (= \log_2 N)$ もよく使われている[5]。ネットワーク上のノードからノードを繋ぐ各経路の平均の情報量を評価する尺度も提案されている [7]。

さて、以下に幾つかの簡単な具体例を取り上げて上記のことを考えてみよう。

例1) 表(Head)および裏(Tail)が出現する確率が等しいコイン投げの時系列

[illegible]

のような時系列が得られたとする(ここで、 H 、 T が等しい確率で出現することは事前には知らされていないとする)。いま、左から右へ向けて進み、長さ v の過去配列、未来配列を記録していく。十分な長い時間、時系列に沿ってそれらを記録した時点で、「次の」未来配列を最適に予測するために、どれだけの過去配列を必要とするのであろうか？ 十分長い時間、時系列を記録し、コイン投げは一切過去の出た目を参照する必要がなく、 T および H が出現する確率は等しいと判定できたとする。

シンボル列の長さ v を 1、2、3、...個にしたときの状態遷移ネットワークを考える。シンボル列“T”および“H”は同じ遷移確率分布を持つ。すなわち、 $P(H|T)=P(H|H)$ 、および $P(T|T)=P(T|H)$ であるため、シンボル列“T”および“H”は同じ因果状態 S_1 に帰属され、因果状態は唯一ひとつ存在することになる。では、 $v=2$ ではどうであろうか？この場合、ありえるシンボル列は“TT”、“TH”、“HH”、“HT”の 4 種類存在するが、すべて同じ遷移確率分布を持ち、状態遷移ネットワークのTopological ComplexityおよびStatistical Complexityが不変であり、過去を参照する必要がないことが容易に確認できる。実は、このコイン投げの例では、過去($v>1$)のみならず、現在、すなわち、いま“T”もしくは“H”であるか否かにも遷移確率分布は依存しないため、 $v=0$ が状態を規定する上で最適・最小な長さということになる(図 2 参照)。このεマシーンにより発生される任意の時系列は与えられたオリジナルの時系列と、統計的に、等価であることが容易に理解できるであろう。

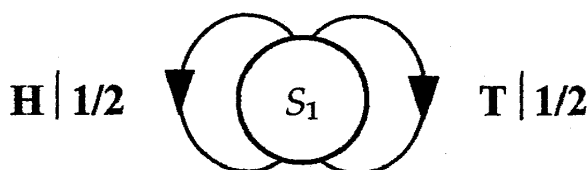


図2 理想的なコイン投げの ε マシン

例 2) 周期 1 の時系列

[illegible]

図 3 すべての同じシンボルから構成される時系列に対するEマシーン

のように、(無限の長さに渡って)すべて同じシンボルが続く周期 1 の時系列を考える。この場合、過去のシンボル列が何であるかに依らずに、未来のシンボル列は常に確率 1で与えられる。この場合も、 $v=0$ が状態を規定する上で最適・最小な長さとなり、因果状態は唯一ひとつ存在することになる(図 3 参照)。

2-4. 階層間の相互相関を取り入れた階層的な状態遷移ネットワーク

最近、我々は生体分子時系列解析に耐えるために、2-3 章における欠点を克服する離散 Wavelet 変換を用いた階層間の相互相関を取り入れた計算力学の手法を新たに開発した[8]。離散 Wavelet 変換は、オリジナルの一次元時系列 $\tau=(\tau_1, \tau_2, \dots, \tau_N)$ を時間・空間スケールに応じて多次元の変数に変換する：

$$\tau = A(n) + D(n) + D(n-1) + \dots + D(1)$$

ここで、 $A(j) = (A_1(j), A_2(j), \dots, A_N(j))$ は τ を 2^j の時間解像度で近似し、 2^j の時間刻みよりも小さい時間スケールの“揺らぎ”を無視し、 $D(j) = (D_1(j), D_2(j), \dots, D_N(j))$ は 2^j の時間解像度における“揺らぎ”を評価することに対応する。また、異なる時間スケールにおける $A(j)$ は

$$A(j-1) = A(j) + D(j) \quad (j \geq 1)$$

を満足する。我々が今回、採用した Haar 基底では、

$$A_i(j) = \left(\sum_{k=i}^{i+2^j-1} \tau_k \right) / 2^j, \quad D_i(j) = \left(\sum_{k=i}^{i+2^{j-1}-1} \tau_k - \sum_{k=i+2^{j-1}}^{i+2^j-1} \tau_k \right) / 2^j \quad (j \geq 1)$$

と表され、 $A(j)$ は τ を 2^j の時間幅での平均化した時系列に、 $D(j)$ は 2^j の時間幅を 2 つに分割して、それぞれの総和の差を時間幅 2^j で割ったものに相当する。図 5 に 3 章で述べるフラビン還元酵素系の電子移動一分子時系列データ $\tau=(\tau_1, \tau_2, \dots, \tau_N)$ を Haar 基底による離散 Wavelet 分解したものを示した(ここでは $n=3$ を例示)。

このようにして多重分解した各成分に対して、2-1、2-2 章で紹介した計算力学を適用し、 ε マシンをそれぞれ構築する。図 6B, 6C に $A(3)$ および $D(3)$ 時系列に対して構成した ε マシン $\varepsilon A(3)$ および $\varepsilon D(3)$ を例示した。 2^3 の時間幅で平均化された $A(3)$ 時系列の ε マシンは、同じ時間スケールにおける平均周りの揺らぎを表す $D(3)$ 時系列のそれに比べて、状態数が多く、より多様であることが分かる。しかしながら、各 ε マシ

ン上を動き回る遷移ダイナミクスはお互いに独立である保証はなく、むしろ、時間領域の近い ε マシン上の遷移ダイナミクスの間には相関が存在するはずである。

$\varepsilon A(3)$ と $\varepsilon D(3)$ のあいだの相関を取り込んだ $\varepsilon A(3), D(3)$ をどのように構成できるであろうか？

さて、 $\varepsilon A(3)$ および $\varepsilon D(3)$ を構成する状態を $S_i A(3)$ ($i=1, \dots, N A(3)$) および $S_j D(3)$ ($j=1, \dots, N D(3)$) と表記しよう(ただし、 $N A(3)$ 、 $N D(3)$ はそれぞれ $\varepsilon A(3)$ 、 $\varepsilon D(3)$ における状態の総数)。ここで、 $\varepsilon A(3)$ と $\varepsilon D(3)$ の状態は、 $A(3)$ および $D(3)$ 時系列の“ある時系列断片”の集合から構成されているので、時間発展に沿って、系が各時刻において $\varepsilon A(3)$ 、 $\varepsilon D(3)$ のどの状態に滞在しているかを同定することができる。そこで $\varepsilon A(3), D(3)$ における状態を $\{S_i A(3), S_j D(3)\}$ の“対”として再定義しよう。簡単のため、その“対”を S_{ij} と略記し、 S_{ij} の出現確率

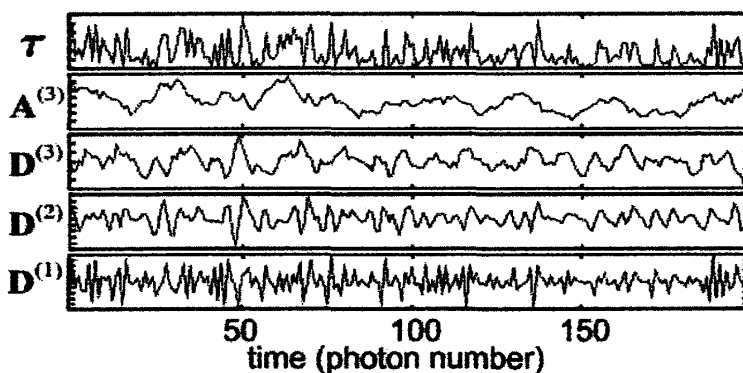
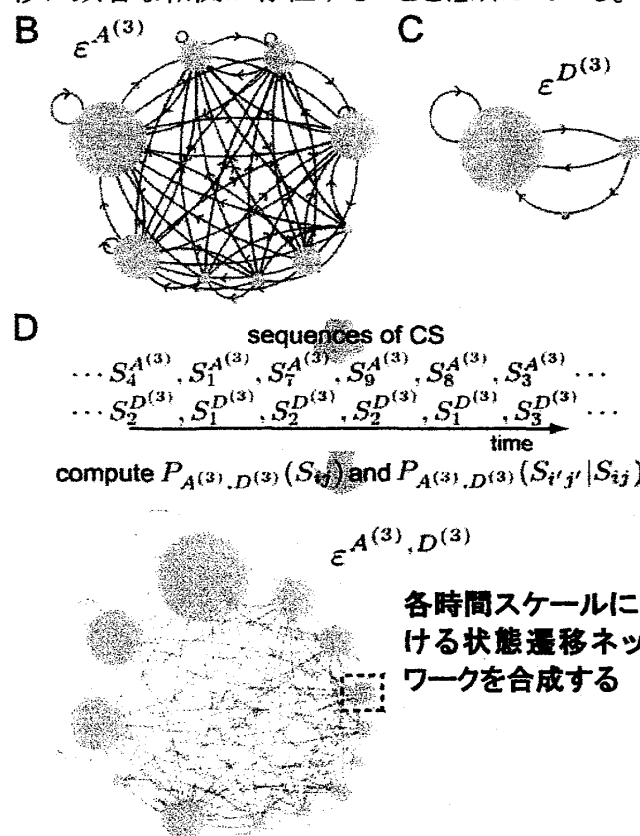


図 5 フラビン還元酵素系の一分子時系列データの離散 Wavelet 分解。

を $P_{A(3),D(3)}(S_{ij})$ と、 S_{ij} から $S_{i'j'}$ への遷移確率を $P_{A(3),D(3)}(S_{ij} | S_{i'j'})$ と書くことにする。 $\varepsilon A(3)$ および $\varepsilon D(3)$ における状態遷移が独立であれば、結合確率 $P_{A(3),D(3)}(S_{ij})$ は独立事象の出現確率の単なる積 $P_{A(3)}(S_i^{A(3)})P_{D(3)}(S_j^{D(3)})$ で表すことができる。一方、異なる ε マシン間の遷移ダイナミクスが相関を持つ場合には、一般に $P_{A(3),D(3)}(S_{ij})$ は $P_{A(3)}(S_i^{A(3)})P_{D(3)}(S_j^{D(3)})$ と異なる。一般に $\varepsilon A(m), \varepsilon D(m)$ は $S_i^{A(m)}$ を平均幅 2^m 内の“揺らぎ”との相関を考慮に入れて、より細かく“篩い分け”したことに相当している。 S_{ij} から $S_{i'j'}$ への遷移確率 $P_{A(3),D(3)}(S_{ij} | S_{i'j'})$ から $\varepsilon A(3), \varepsilon D(3)$ のネットワークを構成することができる。

図 6D に結合された $\varepsilon A(3), \varepsilon D(3)$ のネットワークを示した。 $\varepsilon A(3), \varepsilon D(3)$ における状態数が $\varepsilon A(3)$ および $\varepsilon D(3)$ の状態数 $N_A(3)$ と $N_D(3)$ の積 ($9 \times 3 = 27$) よりも遥かに少ないことが分かる。これは、 $\varepsilon A(3), \varepsilon D(3)$ における状態遷移に顕著な相関が存在することを意味している。時間スケールがより短い $\varepsilon D(2)$ ないし $\varepsilon D(1)$ と $\varepsilon A(3), \varepsilon D(3)$ 上



の状態遷移ダイナミクスが有意な相関をもつ場合には、順次、加えていけばよい。相関が微弱な場合には断熱近似が成立するため、独立に取り扱ってもよいことが期待される。

離散 Wavelet 変換により、一次元時系列データを時空間スケールに依存した時系列成分に分解したことの最も重要な点は、与えられた一次元時系列データのなかから“定常”と見なしえる成分を取り出して、その成分に対して計算力学を適用する点である。離散 Wavelet 変換された各 $A(j), D(j)$ の定常性は自己相関関数やアラン分散[9]を調べることで 2^j の時間領域で定常的であるか否かを容易に確認することができる。一般に $D(j)$ は 2^j の時間領域で近似的に定常である場合が多いのに対し、 $A(j)$ は 2^j 以上の長時間領域のすべての定常・非定常成分を含んでいる。そのため、一般に、 j が大きくなるにつれて(周波数が低くなるほど)、 $A(j)$ のなかのゆっくりした非定常成分が顕在化するため、 $A(j)$ に対する計算力学の適用が困難となる。それゆえ、計算力学を適用するためには、 $A(j), D(j)$ の定常性を評価し、計算力学を適用できる適切なレベル j を見極める必要がある。

図 6 ε マシン $\varepsilon A(3)$ と $\varepsilon D(3)$ の合成: B-C. 約 23ms の時間幅での平均値 $A(3)$ の時系列とその周りの揺らぎ $D(3)$ の時系列のつくる遷移ネットワーク。D 各状態遷移ネットワークを合成して得られる遷移ネットワーク

3. フラビン還元酵素の 1 分子構造揺らぎにおける異常拡散と階層的な状態遷移ネットワーク[8]

図 7 に 1 分子計測されたフラビン還元酵素の 1 分子構造揺らぎ[3a]の自己相関関数の実測値を示した。図から明らかなように、長時間領域 ($> 0.1s$) ではブラウン拡散を呈するのに対し、短時間領域 ($< 0.1s$) では異常拡散を示している。ダイナミクスの非マルコフ性を状態自身の内部構造に押し込めることにより、 ε マシンは状態空間における離散的マルコフ過程と読み替えることができる。そのため、 ε マシン上の運動に由来する物理量は解析的に評価することが可能となる。図中、2 章で紹介した Wavelet 分解された各時間スケールの時系列情報から構成された状態遷移ネットワーク

に基づく理論値(○)を示している。長短の時間領域で実測値を正しく評価していることがわかる。重要なことは、予め実験結果(自己相関関数の形)を再現するように仕組んだモデル構築にはなっていないことである。

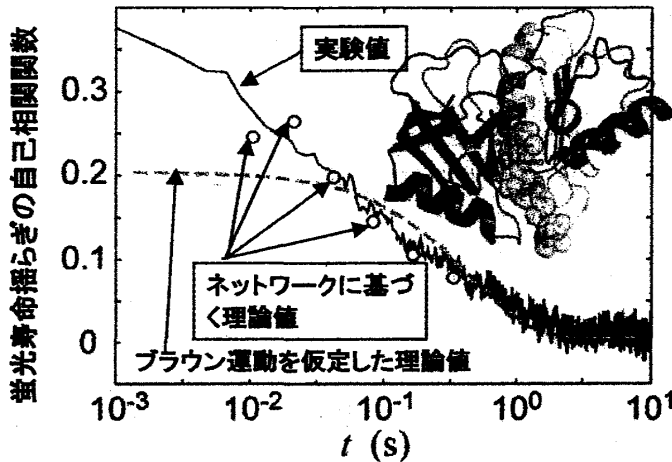


図 7. 蛍光寿命揺らぎの自己相関関数(実験)と階層的状態遷移ネットワークから見積もった自己相関関数(理論)

で表される。

$$Cc(I) = \sum_{J=1} \text{dis}(I:J)P(S_J)$$

ここで、 $\text{dis}(I:J)$ および $P(S_J)$ はそれぞれノードIとノードJの間の“距離”、およびノードJの滞在確率を表し、ノードIとノードJの間の“距離”は状態Iと状態Jに付随する固有の遷移確率分布関数 $P(S_k|S_I)$ と $P(S_k|S_J)$ のあいだのHellinger距離 $d_{HD}(I:J) = \sqrt{\sum_k (\sqrt{P(S_k|S_I)} - \sqrt{P(S_k|S_J)})^2}$ により評価した。すな

わち、(図中の太い両矢印線で示している) $Cc(I)$ の分散が小さい(ネットワークがコンパクトになる)ほど、異なる状態に付随する遷移確率分布関数が一様になってくることを意味している。また、ノード(円)の色合いは、どれくらいの割合の状態と有意に“リンク”しているかを表している(詳細は原著[8]を参照していただきたい)。図8から、

1. ブラウン拡散の時間領域に近づくにつれて、ネットワークがコンパクトになるとともに、ノード(状態)当たりのリンク数が増大し、状態遷移ネットワークのもつ多様性が減少すること、
2. 異常拡散を示す時間領域において、状態間遷移をマルコフ過程と見做すために、ある有限の長さの時系列断片から状態を定義する必要があること、すなわち、局所平衡に基づくマクロな意味のネットワーク描像はこの短時間領域においては成り立たないこと、
3. 通常拡散を呈する時間領域においては、より少数の状態から構成され、「過去の履歴を伴う必要のない、局所平衡概念に基づく通常の準安定状態」として解釈されること、

などを読み取ることができる[8]。このほか紙面の都合上割愛するが、社会学、生態学、WWW、代謝ネットワークなどにおいて重要な概念である“スケールフリー”性[10,11]の存在を解析した。その結果、ブラウン運動で表すことが困難な異常拡散を示す時間領域のネットワークではスケールフリー性を有しているが、通常のブラウン運動として表現される480ms以上の時間領域の時間成分から構成されたネットワークはスケールフリー性を有していないことなどが新たに分かった。スケールフリーであると

いうことは、特徴的な空間スケールが(その時間領域において)存在しないことを意味している。すなわち、ある環境の変化に対し、仮にある特定の長さの経路が遮断されても、他の長さの経路は常に存在することを示唆している。これまで、蛋白質が感じるエネルギー地形は、天然状態に辿るルートにバイアスの掛かった“漏斗(ファネル)”型地形を、自然は進化の所産として獲得してきたものと“トップダウン的に”解釈されてきた。しかしながら、天然蛋白質と人工蛋白質の状態遷移ネットワークではスケールフリー性において有意な差が存在してもおかしくない。一分子時系列情報から直接抽出される“状態遷移ネットワーク構造”のトポロジ的性質を解析することを通して、できるだけ自然な形で背後に存在する“からくり”が解明されていくものと期待される。

また、状態遷移ネットワークの情報から、時間スケールの異なる複数の時系列情報のあいだの情報伝達、情報伝達の異方性、強さを定量化する方法論の開発も始めている。その結果、時間スケールの近接したネットワーク間の相互相関が一般的に大きい、情報伝達の大きさは必ずしも双方向的でなく、遅い時間スケール上のネットワーク上の遷移情報が早い時間スケールにおけるネットワーク上の遷移を支配する傾向があること、などが判明した。

4. 展望

自然科学研究において革命的な発展をもたらすのは、多くの場合、新しい実験技術とその新しい実験事実に基づいた理論・概念の転回である。近年、1 分子計測技術等の飛躍的な進展により、「観測」の在り方が大きな変貌を遂げ、1分子レベルでの生体分子の大規模構造転移、細胞の分化などが直接観測されるようになった。しかしながら“時系列情報から如何にしてその背後に潜むメカニズムに関する知見を抽出しえるのか”は未踏の未解決問題である。本稿で紹介した計算力学は埋め込み論[9]の適用が困難な短い 1 分子時系列情報に対しても、その集合を用いることで、局所平衡を予め仮定しないで、状態および状態間遷移を規定することができる“第一原理的”時系列解析理論といえる。このとき、分子記憶の原理は

- ① 因果状態の内部構造: マルコフ過程とみなしえるためにはどれくらい長いシンボル列(どれくらい長い記憶を保有しえるかに相当する)を必要とするか
- ② ϵ マシンにおける因果状態間の“つながり”の様相: つながりに規則的なパターンが存在するか否か

から探求することができる。例えば、運動の記憶を持たない時間スケールでは、一般的な自由エネルギー地形から類推される状態遷移ネットワーク構造が得られる。しかしながら、記憶を持つ場合には、ダイナミックスの情報を捉えた状態遷移ネットワーク構造に取ってかわることになる。本稿で紹介

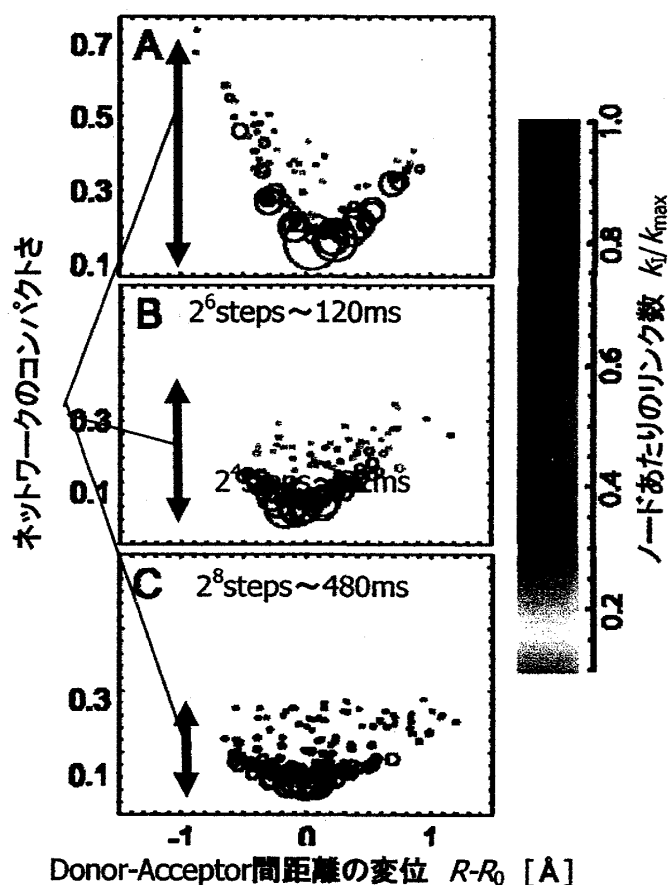


図 8 各時間スケールに応じて変化する状態遷移ネットワーク(理論的には時間無限大の極限で、状態はひとつに収束する)

した(局所平衡仮定を予め必要としない)計算力学に立脚する解析理論は、多次元自由エネルギー地形[12]と比較することを通して、どのような時間スケールで自由エネルギー地形の概念が成り立つか、壊れるか、更には、フラビン還元酵素などにみられたように、生体分子がしばしばその分子の特徴的な時間スケールよりも遙かに長い間なぜ内部記憶を保持しうるのかを解明する上で大きな手掛かりを与えるものと期待される。また、図 8 に示されるように、状態遷移の動態構造は観測する時間スケールに応じて変化する。このような時間スケールに応じて変化する動態描像はまだほとんど理解されていない。一分子時系列解析理論が確立されることで、このような動態構造の存在が一分子観察の時系列情報から次第に解明されていくであろう。更には、1 分子時系列をある長さで断片化することで時間依存の ε マシーンを構成することができる。例えば、環境に適応しながら、系はどのように時々刻々背後に存在する状態空間上のネットワーク構造を変えているかなどを時系列情報から学ぶことも可能となるであろう。ダイナミックスの階層性(断熱近似を越えた階層間の情報伝達)、ネットワークに内在する動的モジュール構造と機能など、興味深いテーマも潜んでいる。

しかしながら克服すべき問題も依然残っている。たとえば、2-2 章の例 1、例 2 においては、因果状態は唯一ひとつ存在し、Topological Complexity および Statistical Complexity とともに 0 である。すなわち、両極端な「完全にランダムな時系列」と「完全に確率 1 で長さ v のシンボル列が生起する時系列」は、ネットワークの複雑さを尺度とする限り、区別することができない。実際の解析対象になる時系列データは、「完全にランダムな時系列」と「完全に確率 1 で長さ v のシンボル列が生起する時系列」のあいだの中間領域に存在する場合がほとんどである。生体分子時系列情報の背後に存在する状態遷移ネットワークのもつ多様性を実際に解析する場合においても、「完全にランダムな時系列」と「完全に確率 1 で長さ v のシンボル列が生起する時系列」に近い状況を同じ組上で比較することはまずない。しかしながら、ランダムネスの中間領域においても、(統計的)複雑さが同じであるにも関わらず、ランダムネスが異なる状況がありえる。時系列情報の背後に存在する生体分子の動態構造を正しく理解するためには、シンボル時系列のランダムネスに相当する指標(例えば、リャプノフ指数、Kolmogorov-Sinai エントロピー、一般には、それらの有限サイズ版である有限サイズリャプノフ指数[13-16]、 (ε, τ) エントロピー[17]、測度エントロピー[5])を複雑さ解析と相補に評価することが必要不可欠となる。

より根本的な問題としては、定常性、符号化の問題がある。計算力学は確率概念に依拠するため、系の定常性を前提としている。そのため非定常性が強い時系列データに対しては、抜本的に方法論を見直すことが求められる。また、計算力学は離散化された時系列データを用いる必要がある。そのため、オリジナルの時系列情報に含まれる情報の損失をできるだけ抑えたシンボル化を行う必要があるが、実用可能な一般的な処方箋は存在しない。

謝辞: 北海道大学電子科学研究所 Chun Biu Li 氏、馬場昭典氏、寺本央氏、清一人氏、カリフォルニア大バークレー校 Haw Yang 氏に感謝する。特に、2-4 章、3 章は Chun Biu Li 氏、Haw Yang 氏との共同研究である。また、いつも有益な議論をしてくれる神戸大学理学部郡司ペギオ幸夫氏、奈良女子大学理学部戸田幹人氏、大阪大学蛋白質研究所高橋聡氏らにもこの場をお借りして感謝したい。本研究は JST/CREST、特定領域「生命システム情報」、「実在系分子理論」、科研費基盤 B 等から財政的支援を受けており、併せてここに感謝する。

参考文献

- [1a] E. Barkai, Y. Jung, and R. Silbey, *Annu. Rev. Phys. Chem.* **55**, 457-507 (2004).
- [1b] X. Zhuang, H. Kim, M. J.B. Pereira, H. P. Babcock, N. G. Walter, and S. Chu, *Science*, **296** 1473(2002)
- [2a] T. Komatsuzaki and R. S. Berry, *Adv. Chem. Phys.* **123** 79 (2002)
- [2b] T. Komatsuzaki and R. S. Berry, *Proc. Natl. Acad. Sci. USA* **98**, 7666 (2001)
- [2c] 小松崎民樹、松永康佑 生物物理 **42**, 285 (2002)

- [3a] H. Yang, G. Luo, P. Karnchanaphanurach, T.-M. Louie, I. Rech, S. Cova, L. Xun, and X. S. Xie, *Science* **302**, 262 (2003)
- [3b] I.V. Gopich, and A. Szabo, *J. Phys. Chem. B.* **107**, 5058-5063 (2003).
- [4] L. Edman and R. Rigler, *Proc. Natl. Acad. Sci. USA*, **97**, 8266(2000)
- [5] C.R. Shalizi and J.P. Crutchfield, *Journal of Statistical Physics*, **104**, 817 (2001)
- [6] L. P. Watkins, H. Yang, *Biophys. J.* **86**, 4015 (2004)
- [7] G. J. Rylance, R. L. Johnston, Y. Matsunaga, C. B. Li, A. Baba, T. Komatsuzaki, *Proc. Natl. Acad. Sci. USA* **103**, 18551 (2006)
- [8] C. B. Li, H. Yang, and T. Komatsuzaki *Proc. Natl. Acad. Sci. USA* **105**, 536-541(2008).
- [9] T. Komatsuzaki, K. Hoshino and Y. Matsunaga, *Adv. Chem. Phys.* **130B**, 257-313 (2005)
- [10] Barabasi *Science* **286**, 509 (1999)
- [11] F. Rao, and A. Caflisch, *J. Mol. Biol.* **342**, 299 (2004).
- [12] A. Baba and T. Komatsuzaki, *Proc. Natl. Acad. Sci. USA* **104**, 19297-19302(2007).
- [13] E. Aurell, G. Boffetta, C. Crisanti, G. Paladin, and A. Vulpiani, *Phys. Rev. Lett.* **77**, 1262 (1996).
- [14] T. Shibata and K. Kaneko, *Phys. Rev. Lett.* **81**, 4116 (1998).
- [15] Y. Matsunaga, C. B. Li, and T. Komatsuzaki, *Phys. Rev. Lett.* **99**, 238103 (2007)
- [16] 松永康佑・小松崎民樹,日本高分子学会誌 高分子 **57**(2),58 (2008)
- [17] P. Gaspard and X.-J. Wang, *Phys. Rep.* **235**, 291 (1993).